ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

European Journal of Combinatorics 27 (2006) 1059–1081

European Journal of Combinatorics

www.elsevier.com/locate/ejc

The number of distinct values in a geometrically distributed sample

Margaret Archibald[a], Arnold Knopfmacher[a], Helmut Prodinger[b]

[a] *The John Knopfmacher Centre for Applicable Analysis and Number Theory, Department of Mathematics, University of the Witwatersrand, P.O. Wits, 2050, Johannesburg, South Africa*

[b] *Department of Mathematics, University of Stellenbosch, 7602, Stellenbosch, South Africa*

Available online 11 July 2006

Abstract

For words of length n, generated by independent geometric random variables, we consider the average and variance of the number of distinct values (=letters) that occur in the word. We then generalise this to the number of values which occur at least b times in the word.

1. Introduction

We consider words $x_1 x_2 \dots x_n$ with letters $x_i \in \{1, 2, \dots\}$. The letter i occurs with (geometric) probability pq^{i-1} where $p + q = 1$, and the letters are considered to be independent, so that $x_1 x_2 \dots x_n$ appears with probability $(p/q)^n q^{x_1+\dots+x_n}$.

The combinatorics of geometric random variables has gained importance because of applications in computer science. We mention just two areas: **skiplists** [3,13,19] and **probabilistic counting** [4,8].

Some of the previous studies relating to combinatorics of geometric random variables are as follows. In [15] the number of left-to-right maxima was investigated in the model of *words* (strings) $a_1 \dots a_n$, where the letters $a_i \in \mathbb{N}$ are independently generated according to the geometric distribution. H.-K. Hwang and his collaborators obtained further results about this limiting behaviour in [2]. The two parameters 'value' and 'position' of the rth left-to-right

E-mail addresses: marchibald@maths.wits.ac.za (M. Archibald), arnoldkn@cam.wits.ac.za (A. Knopfmacher), hproding@sun.ac.za (H. Prodinger).

0195-6698/$ - see front matter
doi:10.1016/j.ejc.2006.06.007

maximum for geometric random variables were considered in a subsequent paper [11]. Other combinatorial questions have been considered [12,14,16,17].

In the paper, we address the following question: How many different letters appear in words of length n, generated by geometric random variables? For this parameter (d_n), we derive expectation and variance. Throughout this paper we use the following notation: $Q = \frac{1}{q}$, $L = \log Q$, $n^* = n(Q-1)$, $\chi_k := \frac{2\pi i k}{L}$ for $k \in \mathbb{Z}$, $k \neq 0$. Also, γ denotes Euler's constant.

Theorem 1. *The number of distinct letters in a word of length n is*

$$\mathbb{E}(d_n) \sim \log_Q n + \frac{\gamma}{L} + \log_Q(Q-1) - \frac{1}{2} + \delta_E(\log_Q n^*), \tag{1}$$

where

$$\delta_E(x) := -\frac{1}{L}\sum_{k\neq 0} \Gamma(-\chi_k) e^{2k\pi i x}.$$

Theorem 2. *The variance for a word of length n is*

$$\mathbb{V}(d_n) \sim \log_Q 2 + \delta_V(\log_Q n^*),$$

where

$$\delta_V(x) := \delta_E(x + \log_Q 2) - \delta_E(x),$$

is a periodic fluctuation with mean zero.

We then generalise this question as follows: How many letters appear at least b times, where $b \geq 1$ is a design parameter.

Theorem 3. *The expected number of digits occurring at least b times in a word is*

$$\mathbb{E}(d_n^{\langle b\rangle}) \sim \log_Q n + \frac{\gamma}{L} + \log_Q(Q-1) - \frac{1}{2} - \frac{1}{L} H_{b-1} + \delta_{E_b}(\log_Q n^*),$$

where

$$\delta_{E_b}(x) := \frac{1}{L}\sum_{j\neq 0} \frac{e^{2\pi i j x}}{\chi_j} \frac{\Gamma(b-\chi_j)}{\Gamma(b)}.$$

Theorem 4. *The variance of this quantity is*

$$\begin{aligned}\mathbb{V}(d_n^{\langle b\rangle}) \sim \log_Q 2 &+ \frac{2}{L}\sum_{i\geq 1} \frac{(-1)^{i+b-1}}{i(Q^i-1)}\binom{i+b-1}{i}\binom{i-1}{b-1} \\ &- \frac{2}{L}\sum_{j=1}^{b-1}\frac{1}{2j}\binom{2j}{j}\sum_{h\geq 0}\binom{-2j}{h}\frac{1}{Q^{h+j}-1} \\ &+ \frac{2}{L}\sum_{h\geq 1}\frac{(-1)^{h-1}}{h(Q^h-1)} - \frac{1}{L}\sum_{j=1}^{b-1}\frac{1}{2j}\binom{2j}{j}2^{-2j} + \delta_{V_b}(\log_Q n^*),\end{aligned}$$

where $\delta_{V_b}(x)$ is a periodic fluctuation with mean zero.

In the asymptotic formulae that we derive, there appear ubiquitous periodic oscillations, due to poles of certain functions at $z = \chi_k$, $k \in \mathbb{Z}$, $k \neq 0$. They are usually tiny, but play an essential role especially in the variance.

2. The distinct value problem

When looking at the number of distinct values in a word, we can make use of exponential generating functions. The total number of letters in the word is represented by n, and k represents the number of distinct values appearing in that word. Our function of interest is the probability generating function

$$F(z, u) := \prod_{i \geq 1} (1 + u(\mathrm{e}^{zpq^{i-1}} - 1)) \tag{2}$$

where the coefficient of $\frac{z^n}{n!}u^k$ is the probability that a word of length n has k distinct values. If letter i occurs at least once, then this will be accommodated by the presence of the u in front of the expression $(\mathrm{e}^{zpq^{i-1}} - 1)$ which represents all non-empty 'sets' of letter i which occur in the word. The problem of the letters appearing at different places in the word is overcome by the use of the exponential generating function.

Note that substituting $u = 1$ into this function gives e^z (since all probabilities sum to 1), which is to be expected because this reduces it to a generating function whose coefficients represent the probability that a word of length n has no restrictions.

Because of the frequent use of Rice's method throughout this paper, we state the following lemma before beginning the proof of Theorem 1 [5,15,21].

Lemma 1. *Let $\mathcal{C}$ be a curve surrounding the points $1, 2, \ldots, n$ in the complex plane, and let $f(z)$ be analytic inside $\mathcal{C}$. Then*

$$\sum_{k=1}^{n} \binom{n}{k} (-1)^k f(k) = -\frac{1}{2\pi i} \int_C [n; z] f(z) \mathrm{d}z,$$

where

$$[n; z] = \frac{(-1)^{n-1} n!}{z(z-1)\cdots(z-n)} = \frac{\Gamma(n+1)\Gamma(-z)}{\Gamma(n+1-z)}. \quad \square$$

By extending the contour of integration, it turns out that under suitable growth conditions (see [5]) the asymptotic expansion of our alternating sum is given by

$$\sum \mathrm{Res}([n; z] f(z)) + \text{smaller order terms},$$

where the sum is taken over all poles different from $1, \ldots, n$. Poles that lie more to the left lead to smaller terms in the asymptotic expansion.

3. The expected value

Let d_n be the number of distinct values in a word of length n, and let $\mathbb{E}(d_n)$ represent the expected value of this quantity. Since [6]

$$\mathbb{E}(d_n) = n![z^n] \frac{\partial}{\partial u} F(z, u) \bigg|_{u=1},$$

and

$$\prod_{i\geq 0}(1+u(e^{zpq^i}-1))\bigg|_{u=1} = \prod_{i\geq 0} e^{zpq^i} = e^{z(p+pq+pq^2+\cdots)} = e^z,$$

we have

$$\begin{aligned}\mathbb{E}(d_n) &= n![z^n]\frac{\partial}{\partial u}\prod_{i\geq 0}(1+u(e^{zpq^i}-1))\bigg|_{u=1}\\ &= n![z^n]e^z\sum_{i\geq 0}\frac{e^{zpq^i}-1}{e^{zpq^i}} = n![z^n]\sum_{i\geq 0}(e^z-e^{z(1-pq^i)})\\ &= \sum_{i\geq 0}(1-(1-pq^i)^n) = \sum_{k=0}^{n}\sum_{i\geq 0}\left(1-\binom{n}{k}(-1)^k p^k q^{ik}\right)\\ &= \sum_{k=1}^{n}\binom{n}{k}(-1)^{k-1}p^k\sum_{i\geq 0}q^{ik} = \sum_{k=1}^{n}\binom{n}{k}(-1)^k\frac{-p^k}{1-q^k}.\end{aligned}$$

However, we cannot easily see what the number of distinct values are from this form. To get a better idea we approximate this alternating sum using Rice's method, by means of Lemma 1.

The first pole we will deal with is at $z=0$, and thus we can approximate our alternating sum (i.e., the expected value) by calculating the residue at $z=0$. We have

$$f(z) = -\frac{(1-Q^{-1})^z}{1-Q^{-z}} = -\frac{(Q-1)^z}{Q^z-1},$$

and from this we can see that there is a double pole at $z=0$ in $[n;z]f(z)$. We thus expand everything to two terms. Firstly, we have [15]:

$$[n;z] = \frac{(-1)^{n-1}n!}{z(z-1)\cdots(z-n)} \sim -\frac{1}{z}(1+zH_n),$$

where $H_n=\sum_{i=1}^{n}\frac{1}{i}$ is the n-th harmonic number. We expand $f(z)$ to get:

$$f(z)\sim -\frac{1}{zL}(1+z\log(Q-1))\left(1-\frac{zL}{2}\right).$$

To calculate the residue at $z=0$ we consider the coefficient of z^{-1} in $[n;z]f(z)$ as $n\to\infty$,

$$[z^{-1}]\frac{1}{z}(1+zH_n)\frac{1}{zL}(1+z\log(Q-1))\left(1-\frac{zL}{2}\right)\sim \log_Q n+\frac{\gamma}{L}+\log_Q(Q-1)-\frac{1}{2},$$

where γ is Euler's constant and the harmonic numbers are approximated by $\log n+\gamma$ as $n\to\infty$. But $f(z) := -\frac{(Q-1)^z}{Q^z-1}$ also has simple poles at $z=\chi_k=\frac{2k\pi i}{L}$, $k\in\mathbb{Z}$, $k\neq 0$. By letting $\varepsilon = z-\chi_k$, we have

$$\begin{aligned}f(z) &= -\frac{(Q-1)^z}{Q^z-1} = -\frac{(Q-1)^{\varepsilon+\chi_k}}{Q^{\varepsilon+\chi_k}-1}\\ &= -\frac{(Q-1)^{\varepsilon}(Q-1)^{\chi_k}}{Q^{\varepsilon}Q^{\chi_k}-1} = (Q-1)^{\chi_k}\left(-\frac{(Q-1)^{\varepsilon}}{Q^{\varepsilon}-1}\right).\end{aligned}$$

Since

$$-\frac{(Q-1)^{\varepsilon}}{Q^{\varepsilon}-1} = -\frac{e^{\varepsilon \log(Q-1)}}{e^{\varepsilon \log Q}-1} \sim -\frac{1}{1+\varepsilon L-1} = -\frac{1}{\varepsilon L},$$

we have that the residue of $f(z)$ is $[\varepsilon^{-1}](-\frac{1}{\varepsilon L}) = -\frac{1}{L}$. From [1], we can see that

$$[n; \chi_k] = \frac{\Gamma(-\chi_k)\Gamma(n+1)}{\Gamma(n+1-\chi_k)} \sim \Gamma(-\chi_k) n^{\chi_k},$$

and

$$(Q-1)^{\chi_k} n^{\chi_k} = e^{(\log n^*)\chi_k} = e^{2k\pi i \log_Q n^*},$$

which means that we can write the small fluctuations as $\delta_E(\log_Q n^*)$, with

$$\delta_E(x) = -\frac{1}{L}\sum_{k\neq 0} \Gamma(-\chi_k) e^{2k\pi i x}.$$

This concludes the proof of Theorem 1. ■

Remark. It is of interest to compare this result with the mean of the *largest value* in a geometrically distributed sample of n letters, denoted by $\mathbb{E}(M_n)$, due to Szpankowski and Rego [22]:

$$\mathbb{E}(M_n) \sim \log_Q n + \frac{\gamma}{L} + \frac{1}{2} + \delta_E(\log_Q n). \tag{3}$$

Ignoring the small fluctuating terms we see that the expected *number of missing values* in the range 1 up to $\mathbb{E}(M_n)$ is asymptotically given by

$$\mathbb{E}(M_n) - \mathbb{E}(d_n) \approx 1 - \log_Q(Q-1). \tag{4}$$

Observe that as Q goes from 1 to ∞, (4) goes monotonically from infinity to 0. Thus $\mathbb{E}(d_n) \to \mathbb{E}(M_n)$ as $q = Q^{-1} \to 0$, which is intuitively clear, since the limiting word is just the sequence **111 . . . 1** with only one distinct value.

Our expected value is sandwiched between (3) and the number of consecutive non-empty boxes (equivalently the first value which does not occur in our sample). The case $q = \frac{1}{2}$ is dealt with in [4], where this value was given as

$$\mathbb{E}(c_n) \sim \log_2 n + \log_2 \varphi + P(\log_2 n)$$

for $\varphi = 0.77351\ldots$ and a periodic function $P(x)$ with period 1 and amplitude bounded by 10^{-5}. We calculate the constants numerically for the case $Q = 2$ to see by how much each expected value differs from the next. The constants which determine the ordering $\mathbb{E}(c_n) \le \mathbb{E}(d_n) \le \mathbb{E}(M_n)$ are (respectively, to four decimal places) -0.3705; 0.3327; 1.3327.

4. The variance

The formula for variance from a generating function is [6]

$$\mathbb{V} = n![z^n]\frac{\partial^2}{\partial u^2} F(z,u)\bigg|_{u=1} + \mathbb{E}(d_n) - \mathbb{E}^2(d_n). \tag{5}$$

Using the generating function in (2) we can calculate the first term of the variance as follows: Let $f_i(z,u) := 1 + u(\mathrm{e}^{zpq^i} - 1)$, then

$$\begin{aligned} n![z^n]\frac{\partial^2}{\partial u^2}F(z,u)\bigg|_{u=1} &= n![z^n]\frac{\partial^2}{\partial u^2}\prod_{i\ge1} f_i(z,u)\bigg|_{u=1} \\ &= n![z^n]\prod_{i\ge1} f_i(z,u)2\sum_{j<k}\frac{\frac{\partial}{\partial u}f_j(z,u)}{f_j(z,u)}\cdot\frac{\frac{\partial}{\partial u}f_k(z,u)}{f_k(z,u)} + \prod_{i\ge1} f_i(z,u)\sum_j \frac{\frac{\partial^2}{\partial u^2}f_j(z,u)}{f_j(z,u)}\Bigg|_{u=1} \\ &= n![z^n]2\mathrm{e}^z\sum_{j<k}\frac{\mathrm{e}^{zpq^j}-1}{\mathrm{e}^{zpq^j}}\cdot\frac{\mathrm{e}^{zpq^k}-1}{\mathrm{e}^{zpq^k}} \\ &= n![z^n]2\sum_{j<k}(\mathrm{e}^z - \mathrm{e}^{z(1-pq^j)} - \mathrm{e}^{z(1-pq^k)} + \mathrm{e}^{z(1-pq^j-pq^k)}) \\ &= 2\sum_{j<k}(1-(1-pq^j)^n - (1-pq^k)^n + (1-pq^j-pq^k)^n). \end{aligned}$$

This quantity can be split up (preserving convergence) as follows in order to be dealt with in two parts:

$$\begin{aligned} n![z^n]\frac{\partial^2}{\partial u^2}F(z,u)\bigg|_{u=1} &= 2\sum_{j<k}[1-(1-pq^k)^n] \\ &\quad + 2\sum_{j<k}[(1-pq^j-pq^k)^n - (1-pq^j)^n]. \end{aligned}$$

The reason for this is that now the summand of the first sum is independent of j and can be dealt with separately from the second sum which requires a slightly different approach. The factor of two is temporarily ignored.

Part (i): Since $1-(1-pq^k)^n$ is independent of j,

$$\begin{aligned} \sum_{k\ge0}\sum_{j=0}^{k-1}[1-(1-pq^k)^n] = \sum_{k\ge0}k[1-(1-pq^k)^n] &= \sum_{k\ge0}k\left[-\sum_{i=1}^{n}\binom{n}{i}(-pq^k)^i\right] \\ &= \sum_{i=1}^{n}\binom{n}{i}(-1)^i\frac{-(Q-1)^i}{(Q^i-1)^2}. \end{aligned}$$

So $f(z) = \frac{-(Q-1)^z}{(Q^z-1)^2}$ and we have a triple pole at $z=0$ as $[n;z]$ has a simple pole and $\frac{-(Q-1)^z}{(Q^z-1)^2}$ has a double pole. To use Rice's method we expand to three terms and get [15]

$$\frac{(-1)^{n-1}n!}{z(z-1)\cdots(z-n)} \sim -\frac{1}{z}\left(1 + zH_n + z^2\frac{H_n^2 + H_n^{(2)}}{2}\right),$$

and

$$-\frac{(Q-1)^z}{(Q^z-1)^2} \sim -\frac{1}{z^2L^2}\left(1 + z\log(Q-1) + \frac{z^2\log^2(Q-1)}{2}\right)\left(1 - zL + \frac{5z^2L^2}{12}\right).$$

We now briefly note that as $n \to \infty$ [20, page 187]

$$H_n^2 \sim (\log n + \gamma)^2 = \log^2 n + 2\gamma \log n + \gamma^2 \quad \text{and} \quad H_n^{(2)} \sim \frac{\pi^2}{6}.$$

The residue for the triple pole at $z = 0$ is

$$\begin{aligned}
&[z^{-1}]\frac{1}{z^3L^2}\left(1 + z\log(Q-1) + \frac{z^2\log^2(Q-1)}{2}\right)\left(1 - zL + \frac{5z^2L^2}{12}\right)\\
&\quad\times\left(1 + zH_n + z^2\frac{H_n^2 + H_n^{(2)}}{2}\right)\\
&\quad= \frac{\log_Q^2(Q-1)}{2} + \frac{5}{12} + \frac{H_n^2 + H_n^{(2)}}{2L^2} - \log_Q(Q-1) + \frac{\log_Q(Q-1)H_n}{L} - \frac{H_n}{L}\\
&\quad\sim \frac{1}{2}\log_Q^2 n + \frac{\gamma}{L}\log_Q n + \log_Q(Q-1)\log_Q n - \log_Q n + \frac{1}{2}\log_Q^2(Q-1)\\
&\qquad - \log_Q(Q-1) + \frac{\gamma}{L}\log_Q(Q-1) + \frac{5}{12} + \frac{\pi^2}{12L^2} + \frac{\gamma^2}{2L^2} - \frac{\gamma}{L} \quad (\text{as } n \to \infty).
\end{aligned}$$

Now $f(z)$ also has double poles at $z = \chi_k,\ k \neq 0$. By letting $\varepsilon = z - \chi_k$, we can use results from the expected value to get

$$f(z) = \frac{-(Q-1)^{\chi_k}(Q-1)^{\varepsilon}}{(Q^{\varepsilon}Q^{\chi_k} - 1)^2} = (Q-1)^{\chi_k}\left(-\frac{(Q-1)^{\varepsilon}}{(Q^{\varepsilon}-1)^2}\right).$$

We have already expanded the fraction to three terms, so we merely note the expansion to two terms as

$$-\frac{(Q-1)^{\varepsilon}}{(Q^{\varepsilon}-1)^2} \sim \frac{-1}{\varepsilon^2L^2}(1 + \varepsilon\log(Q-1))(1 - \varepsilon L).$$

Lastly,

$$\Lambda := [n; \chi_k] = \frac{\Gamma(-z)\Gamma(n+1)}{\Gamma(n+1-z)}$$

needs to be expanded to 2 terms around $z = \chi_k$. Using a Taylor expansion we can write

$$\Gamma(-z) \sim \Gamma(-\chi_k) - \Gamma'(-\chi_k)(z - \chi_k) = \Gamma(-\chi_k)\left(1 - \psi(-\chi_k)(z - \chi_k)\right)$$

and similarly

$$\Gamma(n+1-z) \sim \Gamma(n+1-\chi_k)(1 - \psi(n+1-\chi_k)(z - \chi_k)).$$

This means that with the same substitution as before ($\varepsilon = z - \chi_k$), we have

$$\Lambda \sim \Gamma(n+1)\frac{\Gamma(-\chi_k)}{\Gamma(n+1-\chi_k)}[1 - \psi(-\chi_k)\varepsilon + \psi(n+1-\chi_k)\varepsilon].$$

We approximate the ψ function by [1, page 259]

$$\psi(n+1-\chi_k) \sim \log(n+1-\chi_k) = \log\left(n\left(1 + \frac{1-\chi_k}{n}\right)\right) \sim \log n \quad \text{as } n \to \infty,$$

so that

$$\begin{aligned}
\Lambda &\sim \Gamma(n+1)\frac{\Gamma(-\chi_k)}{\Gamma(n+1-\chi_k)}\left[1-\psi(-\chi_k)\varepsilon+\varepsilon\log n\right]\\
&= \Gamma(-\chi_k)\frac{\Gamma(n+1)}{\Gamma(n+1-\chi_k)}\left[1-\psi(-\chi_k)\varepsilon+\varepsilon\log n\right]\\
&\sim \Gamma(-\chi_k)n^{\chi_k}\left[1-\psi(-\chi_k)\varepsilon+\varepsilon\log n\right] \quad \text{as } n\to\infty.
\end{aligned}$$

If we put this together with the expansion for $f(z)$ obtained above we get

$$\begin{aligned}
\Lambda f(z) \sim{}& (Q-1)^{\chi_k}\frac{-1}{\varepsilon^2L^2}\Gamma(-\chi_k)n^{\chi_k}\left[1-\psi(-\chi_k)\varepsilon+\varepsilon\log n\right]\\
&\times(1+\varepsilon\log(Q-1))(1-\varepsilon L) \quad \text{as } n\to\infty,
\end{aligned}$$

and by rewriting $(Q-1)^{\chi_k}n^{\chi_k}$ as $\mathrm{e}^{\chi_k\log n^*}=\mathrm{e}^{2\pi ik\log_Q n^*}$, we get the residue from the poles at $z=\chi_k$, $k\neq 0$ to be

$$\begin{aligned}
&\sum_{k\neq 0}\mathrm{e}^{2\pi ik\log_Q n^*}\Gamma(-\chi_k)\frac{-1}{L^2}\left[-\psi(-\chi_k)+\log n+\log(Q-1)-L\right]\\
&\quad= \frac{-1}{L}\sum_{k\neq 0}\mathrm{e}^{2\pi ik\log_Q n^*}\Gamma(-\chi_k)\left[\log_Q n-\frac{\psi(-\chi_k)}{L}+\log_Q(Q-1)-1\right].
\end{aligned}$$

So the total result for part (i) is:

$$\begin{aligned}
&\sum_{k\geq 0}\sum_{j=0}^{k-1}[1-(1-pq^k)^n]\\
&\sim \frac{1}{2}\log_Q^2 n+\frac{\gamma}{L}\log_Q n+\log_Q(Q-1)\log_Q n-\log_Q n+\frac{1}{2}\log_Q^2(Q-1)\\
&\quad-\log_Q(Q-1)+\frac{\gamma}{L}\log_Q(Q-1)+\frac{5}{12}+\frac{\pi^2}{12L^2}+\frac{\gamma^2}{2L^2}-\frac{\gamma}{L}\\
&\quad-\frac{1}{L}\sum_{k\neq 0}\mathrm{e}^{2\pi ik\log_Q n^*}\Gamma(-\chi_k)\left[\log_Q n-\frac{\psi(-\chi_k)}{L}+\log_Q(Q-1)-1\right].
\end{aligned} \tag{6}$$

Part (ii): Applying the binomial theorem gives

$$\sum_{j<k}[(1-pq^j-pq^k)^n-(1-pq^j)^n]=\sum_{i=1}^{n}\binom{n}{i}(-1)^{i-1}\sum_{j<k}[(pq^j)^i-(pq^j+pq^k)^i]$$

(since term $i=0$ is zero). This is now written in the correct form for Rice's method to be used, where (for $m=k-j$)

$$\begin{aligned}
f(z) &= -\sum_{j<k}[(pq^j)^z-(pq^j+pq^k)^z]\\
&= -\sum_{j\geq 0}(pq^j)^z\sum_{m\geq 1}[1-(1+q^m)^z] =: -\frac{p^z}{1-q^z}\,g(z).
\end{aligned}$$

We now expand $g(z)$ around $z = 0$

$$\begin{aligned} g(z) &= \sum_{m\geq 1}[1-\mathrm{e}^{z\log(1+q^m)}] = \sum_{m\geq 1}\left[-z\log(1+q^m) - \frac{z^2\log^2(1+q^m)}{2} + \cdots\right] \\ &= z\sum_{m\geq 1}\sum_{k\geq 1}\frac{(-1)^k(q^m)^k}{k} - \frac{z^2}{2}\sum_{m\geq 1}\left(\sum_{k\geq 1}\frac{(-1)^{k+1}(q^m)^k}{k}\right)^2 + \cdots \\ &= z\sum_{k\geq 1}\frac{(-1)^k}{k}\frac{q^k}{1-q^k} - \frac{z^2}{2}\sum_{k\geq 1}\sum_{j\geq 1}\frac{(-1)^{k+j}}{kj}\frac{q^{k+j}}{1-q^{k+j}} + \cdots. \end{aligned}$$

Now, these sums can be evaluated by Mathematica to give constants, for example, if $q = 1/2$ then

$$\alpha := \sum_{k\geq 1}\frac{(-1)^k}{k}\frac{q^k}{1-q^k} = -0.868877 \quad \text{and}$$

$$\beta := -\frac{1}{2}\sum_{k\geq 1}\sum_{j\geq 1}\frac{(-1)^{k+j+2}}{kj}\frac{q^{k+j}}{1-q^{k+j}} = -0.116506.$$

So $g(z)$ can be written as $g(z) = \alpha z + \beta z^2 + \cdots$. Thus when we use Rice's method we have a simple pole at $z = 0$. Consequently we expand everything to one term, giving

$$\frac{(-1)^{n-1}n!}{z(z-1)\cdots(z-n)} \sim -\frac{1}{z} \quad \text{and} \quad -\frac{(Q-1)^z}{Q^z-1} \sim -\frac{1}{zL},$$

and $g(z) \sim \alpha z$. The residue for $z = 0$ is thus $[z^{-1}]\left(-\frac{1}{z}\right)\left(-\frac{1}{zL}\right)\alpha z = \frac{\alpha}{L}$. But $-\frac{(Q-1)^z}{Q^z-1}$ also has simple poles at $z = \chi_k$, $k \neq 0$. To see how $g(z)$ behaves around χ_k, we rearrange,

$$g(\chi_k) = \sum_{m\geq 1}\left[1-(1+q^m)^{\chi_k}\right] = -\sum_{m\geq 1}\sum_{l\geq 1}\binom{\chi_k}{l}(q^m)^l = -\sum_{l\geq 1}\binom{\chi_k}{l}\frac{1}{Q^l-1}.$$

This is the contribution of $g(z)$. The residue of $-\frac{(Q-1)^z}{Q^z-1}$ was dealt with in the expected value section, we again let $\varepsilon = z - \chi_k$, and get:

$$\frac{(Q-1)^{\varepsilon}}{Q^{\varepsilon}-1} \sim \frac{1}{\varepsilon L},$$

and so the residue is $\frac{1}{L}$. As in part (i), $(Q-1)^{\chi_k}[n;\chi_k] \sim \Gamma(-\chi_k)\mathrm{e}^{2\pi ik\log_Q n^*}$, and so the contribution from the simple poles at $z = \chi_k$ is

$$-\frac{1}{L}\sum_{k\neq 0} g(\chi_k)\Gamma(-\chi_k)\mathrm{e}^{2\pi ik\log_Q n^*},$$

which means that the total result for part (ii) is

$$\sum_{j<k}[(1-pq^j-pq^k)^n - (1-pq^j)^n] = \frac{\alpha}{L} - \frac{1}{L}\sum_{k\neq 0} g(\chi_k)\Gamma(-\chi_k)\mathrm{e}^{2k\pi i\log_Q n^*}.$$

To compute the variance we also need (1) and

$$\begin{aligned}\mathbb{E}^2(d_n) \sim\ & \log_Q^2 n + 2\log_Q n\,\delta_E(\log_Q n^*) + 2\log_Q n \log_Q(Q-1)\\ & + \frac{2\gamma \log_Q n}{L} - \log_Q n + \frac{1}{4} - \frac{\gamma}{L} + \frac{\gamma^2}{L^2}\\ & - \log_Q(Q-1) + \frac{2\gamma \log_Q(Q-1)}{L} + \log_Q^2(Q-1)\\ & + 2\log_Q(Q-1)\delta_E(\log_Q n^*) - \delta_E(\log_Q n^*)\\ & + \frac{2\gamma\delta_E(\log_Q n^*)}{L} + \delta_E^2(\log_Q n^*).\end{aligned}$$

We can now put all of these together (remembering that part (i) and part (ii) must include a factor of two) to get

$$\begin{aligned}\mathbb{V}(d_n) =\ & n![z^n]\frac{\partial^2}{\partial u^2}F(z,u)\Big|_{u=1} + \mathbb{E}(d_n) - \mathbb{E}^2(d_n)\\ \sim\ & \frac{1}{12} + \frac{\pi^2}{6L^2} + \frac{2\alpha}{L} - \frac{2}{L}\sum_{k\neq 0} \mathrm{e}^{2\pi i k \log_Q n^*}\Gamma(-\chi_k)\\ & \times\left[\log_Q n - \frac{\psi(-\chi_k)}{L} + \log_Q(Q-1) - 1\right]\\ & - \frac{2}{L}\sum_{k\neq 0} g(\chi_k)\Gamma(-\chi_k)\mathrm{e}^{2\pi i k \log_Q n^*} + \delta_E(\log_Q n^*) - 2\log_Q n\delta_E(\log_Q n^*)\\ & - 2\log_Q(Q-1)\delta_E(\log_Q n^*) + \delta_E(\log_Q n^*) - \frac{2\gamma\delta_E(\log_Q n^*)}{L} - \delta_E^2(\log_Q n^*).\end{aligned}$$

We can split up the $\delta_E^2(\log_Q n^*)$ term into a constant term (the mean of the fluctuating function) and a fluctuating function of mean zero. Let (see [9])

$$\delta_E^2(x) = [\delta_E^2]_0 + \hat{\delta}_E(x) = \frac{\pi^2}{6L^2} + \frac{1}{12} - \log_Q 2 - \frac{2}{L}\sum_{h\geq 1}\frac{(-1)^{h-1}}{h(Q^h-1)} + \hat{\delta}_E(x)$$

where $\hat{\delta}_E(x) = \frac{1}{L^2}\sum_{k\neq 0}\sum_{j\neq 0,\neq k}\Gamma(-\chi_j)\Gamma(-\chi_{k-j})\mathrm{e}^{2\pi i k x}$. Then we have (for $\alpha = \sum_{k\geq 1}\frac{(-1)^k}{k}\frac{q^k}{1-q^k}$)

$$\begin{aligned}\mathbb{V}(d_n) &\sim \log_Q 2 + \frac{2\alpha}{L} + \frac{2}{L}\sum_{h\geq 1}\frac{(-1)^{h-1}}{h(Q^h-1)} + \delta_V(\log_Q n^*)\\ &= \log_Q 2 + \delta_V(\log_Q n^*),\end{aligned}$$

where

$$\begin{aligned}\delta_V(x) &= \frac{2}{L}\sum_{k\neq 0}\Gamma(-\chi_k)\mathrm{e}^{2k\pi i x}\left[\frac{\psi(-\chi_k)}{L} - g(\chi_k) + \frac{\gamma}{L}\right] - \hat{\delta}_E(x)\\ &= \delta_E(x+\log_Q 2) - \delta_E(x),\end{aligned} \tag{7}$$

with $g(x) = -\sum_{l\geq 1}\binom{x}{l}\frac{q^l}{1-q^l}$. Appendix A provides the simplifications for (7). This concludes the proof of Theorem 2. ∎

Extreme cases of α

For interest we look at the extreme cases of α in $g(z) = \alpha z + \beta z^2 + \cdots$. As $q \to 0$, $\alpha \to 0$. If $q \to 1$, then by letting $q = \mathrm{e}^{-t}$, we can consider $t \to 0$. Rewriting, we have the quantity

$$\sum_{k\geq 1} \frac{(-1)^k}{k} \frac{\mathrm{e}^{-tk}}{1-\mathrm{e}^{-tk}} = -\sum_{k\geq 1} \frac{(-1)^{k-1}}{k} \frac{1}{\mathrm{e}^{tk}-1}.$$

This can be found in the appendix of [10], and by calling it $g(t)$ we get the following result from that paper, which makes use of Mellin transforms to get:

$$\alpha = g(t) = -\frac{\pi^2}{12t} + \frac{\log 2}{2} - \frac{t}{24} + g\left(\frac{2\pi^2}{t}\right).$$

This identity holds for $0 < t < 2\pi^2$. We are interested in what happens as $t \to 0$. Since

$$g(t) = -\sum_{k\geq 1} \frac{(-1)^{k-1}}{k} \frac{1}{\mathrm{e}^{tk}-1},$$

it can be seen that $g(\frac{2\pi^2}{t}) \to 0$ as $t \to 0$, and thus the last term in the expression for α is small enough to be insignificant. The remaining three terms provide an approximation for α near $q = 1$ where $t = \log \frac{1}{q}$.

5. General case: Number of letters occurring at least b times

We now generalise to the case where we consider the number of values in a word which appear at least b times. Our probability generating function needs to be extended to

$$F_b(z,u) := \prod_{i\geq 0} \left(\sum_{k=0}^{b-1} \frac{(zpq^i)^k}{k!} + u \left(\mathrm{e}^{zpq^i} - \sum_{k=0}^{b-1} \frac{(zpq^i)^k}{k!} \right) \right).$$

6. The expected value (general case)

$$\begin{aligned}
\mathbb{E}(d_n^{\langle b\rangle}) &= n![z^n] \frac{\partial}{\partial u} F_b(z,u)\Big|_{u=1} = n![z^n] \sum_{i\geq 0} \frac{\mathrm{e}^z \left(\mathrm{e}^{zpq^i} - \sum_{k=0}^{b-1} \frac{(zpq^i)^k}{k!} \right)}{\mathrm{e}^{zpq^i}} \\
&= \sum_{i\geq 0} \left(1 - \sum_{k=0}^{b-1} \binom{n}{k} (1-pq^i)^{n-k} (pq^i)^k \right) \\
&= \sum_{i\geq 0} \left(1 - \sum_{j\geq 0} \binom{n}{j} (-pq^i)^j \right) - \sum_{i\geq 0} \sum_{k=1}^{b-1} \binom{n}{k} \sum_{j\geq 0} \binom{n-k}{j} (-pq^i)^j (pq^i)^k \\
&= \sum_{j=1}^{n} \binom{n}{j} (-1)^{j-1} \frac{p^j}{1-q^j} - \sum_{k=1}^{b-1} \binom{n}{k} \sum_{j=0}^{n-k} \binom{n-k}{j} (-1)^j \frac{p^{j+k}}{1-q^{j+k}}.
\end{aligned}$$

The first term is our original expected value (i.e., for the number of distinct values) and Rice's method (with the contour of integration surrounding $0, \ldots, n$) can be used for the inner sum of the second term. The function $f_k(z) = \frac{(Q-1)^{z+k}}{Q^{z+k}-1}$ has a pole at $z = -k$, and for $\varepsilon = z + k$ we have

$$f_k(z) \sim \frac{(Q-1)^{\varepsilon}}{Q^{\varepsilon}-1} = \frac{1}{\varepsilon \log Q} \quad \text{as } \varepsilon \to 0$$

i.e., the residue is $\frac{1}{L}$. The contribution of $[N; z]$ (where $N = n - k$) around $z = -k$ is

$$[n-k; -k] = \frac{(-1)^{n-k-1}(n-k)!}{(-1)^{n-k+1}(k)(k+1)\cdots(n)} = \frac{(n-k)!(k-1)!}{n!},$$

and so the total residue is $\frac{(n-k)!(k-1)!}{Ln!}$. This can now be substituted in as the inner sum, giving

$$\sum_{k=1}^{b-1} \binom{n}{k} \frac{(n-k)!(k-1)!}{Ln!} = \sum_{k=1}^{b-1} \frac{1}{Lk} = \frac{1}{L} H_{b-1}.$$

Lastly, we need to calculate the fluctuations contributed by the simple poles at $z+k = \chi_j$, $j \in \mathbb{Z}$, $j \neq 0$. We have $f_k(z) = \frac{(Q-1)^{z+k}}{Q^{z+k}-1}$ and let $\varepsilon = z + k - \chi_j$, then

$$f_k(z) = \frac{(Q-1)^{\varepsilon+\chi_j}}{Q^{\varepsilon+\chi_j}-1} = (Q-1)^{\chi_j} \frac{(Q-1)^{\varepsilon}}{Q^{\varepsilon}-1} \sim (Q-1)^{\chi_j} \frac{1}{\varepsilon L}, \tag{8}$$

as in the previous expected value, so the residue is $(Q-1)^{\chi_j} \frac{1}{L}$. The contribution of $[N; z]$ around $z = -k + \chi_j$ is [1]

$$\begin{aligned}[n-k; -k+\chi_j] &= \frac{\Gamma(k-\chi_j)\Gamma(n-k+1)}{\Gamma(n-k+1+k-\chi_j)} = \frac{\Gamma(k-\chi_j)\Gamma(n-k+1)}{\Gamma(n+1-\chi_j)} \\ &\sim \Gamma(k-\chi_j) n^{\chi_j - k}. \end{aligned} \tag{9}$$

Again we can write $(Q-1)^{\chi_j} n^{\chi_j} = e^{2\pi i j \log_Q n^*}$. For each value of k we have a contribution of $\frac{1}{L} \sum_{j\neq 0} \Gamma(k-\chi_j) n^{-k} e^{2\pi i j x}$. We sum this to get

$$\sum_{k=1}^{b-1} \binom{n}{k} \frac{1}{L} \sum_{j \neq 0} \Gamma(k-\chi_j) n^{-k} e^{2\pi i j x},$$

which we can subtract from the δ function in the case $b = 1$. Thus as $n \to \infty$,

$$\begin{aligned} &\frac{-1}{L} \sum_{j\neq 0} e^{2\pi i j x} \left[\Gamma(-\chi_j) + \sum_{k=1}^{b-1} \binom{n}{k} \Gamma(k-\chi_j) n^{-k} \right] \\ &= \frac{-1}{L} \sum_{k=0}^{b-1} \frac{n^{\underline{k}}}{k! n^k} \sum_{j \neq 0} e^{2\pi i j x} \Gamma(k-\chi_j) \\ &\sim \frac{-1}{L} \sum_{k=0}^{b-1} \sum_{j\neq 0} e^{2\pi i j x} \frac{\Gamma(k-\chi_j)}{k!} = \frac{1}{L} \sum_{j \neq 0} \frac{e^{2\pi i j x}}{\chi_j} \frac{\Gamma(b-\chi_j)}{\Gamma(b)}. \end{aligned}$$

Thus the expected number of digits occurring at least b times in a word is

$$\mathbb{E}(d_n^{\langle b\rangle}) \sim \log_Q n + \frac{\gamma}{L} + \log_Q(Q-1) - \frac{1}{2} - \frac{1}{L}H_{b-1} + \delta_{E_b}(\log_Q n^*),$$

where $\delta_{E_b}(x) = \frac{1}{L}\sum_{j\neq 0} \frac{e^{2\pi ijx}}{\chi_j}\frac{\Gamma(b-\chi_j)}{\Gamma(b)}$. This concludes the proof of Theorem 3. ■

7. The variance (general case)

The corresponding second factorial moment in (5) can be calculated as follows (bearing in mind that all double partial derivatives with respect to u are 0, as each term is linear with respect to u),

$$\begin{aligned}
\frac{\partial^2}{\partial u^2}F_b(z,u)\Big|_{u=1} &= 2e^z \sum_{0\le l<j} \frac{\left(e^{zpq^l} - \sum_{k=0}^{b-1}\frac{(zpq^l)^k}{k!}\right)\left(e^{zpq^j} - \sum_{k=0}^{b-1}\frac{(zpq^j)^k}{k!}\right)}{e^{zpq^l}e^{zpq^j}}\\
&= 2\sum_{0\le l<j}\Bigg[e^z - e^{z(1-pq^l)}\left(1+\cdots+\frac{(zpq^l)^{b-1}}{(b-1)!}\right)\\
&\quad - e^{z(1-pq^j)}\left(1+\cdots+\frac{(zpq^j)^{b-1}}{(b-1)!}\right)\\
&\quad + e^{z(1-pq^l-pq^j)}\left(1+\cdots+\frac{(zpq^l)^{b-1}}{(b-1)!}\right)\left(1+\cdots+\frac{(zpq^j)^{b-1}}{(b-1)!}\right)\Bigg].
\end{aligned} \tag{10}$$

Terms one and three of (10) can be combined to give $2\sum_{j\ge 0} j[e^z - e^{z(1-pq^j)}(1+\cdots+\frac{(zpq^j)^{b-1}}{(b-1)!})]$, with coefficients:

$$2\sum_{j\ge 0} j\left[1-(1-pq^j)^n - \cdots - \binom{n}{b-1}(pq^j)^{b-1}(1-pq^j)^{n-(b-1)}\right].$$

Now this can be split up into $2\sum_{j\ge 0} j\left[1-(1-pq^j)^n\right]$ (which is known from (6), and

$$-2\sum_{j\ge 0} j\left[npq^j(1-pq^j)^{n-1} + \cdots + \binom{n}{b-1}(pq^j)^{b-1}(1-pq^j)^{n-(b-1)}\right].$$

A typical term is:

$$\begin{aligned}
&-2\binom{n}{s}\sum_{j\ge 0} j\,(pq^j)^s(1-pq^j)^{n-s}\\
&\quad = -2\binom{n}{s}\sum_{h=0}^{N}\binom{N}{h}(-1)^h\frac{(pq)^{h+s}}{(1-q^{h+s})^2} \qquad (N := n-s),
\end{aligned}$$

for which there is a double pole at $z = -s$. Again Rice's method can be used. Let $\varepsilon = z + s$. Then we have to expand around $\varepsilon \to 0$ to two terms:

$$f(z) := \frac{(pq)^{z+s}}{(1-q^{z+s})^2} = \frac{(pq)^{\varepsilon}}{(1-q^{\varepsilon})^2} = \frac{(Q-1)^{\varepsilon}}{(Q^{\varepsilon}-1)^2} \sim \frac{1}{\varepsilon^2 L^2}\left(1+\varepsilon\log(Q-1) - \varepsilon L\right).$$

The Taylor expansion of $[N; z]$ around $z = -s$ (i.e., around $\varepsilon = 0$) to two places (this can be done by Mathematica) is

$$[n-s; -s] \sim \frac{(n-s)!(s-1)!}{n!}[1 + \varepsilon\,\psi(n+1) - \varepsilon\,\psi(s)].$$

We get the residue by multiplying these two expansions:

$$\frac{(n-s)!(s-1)!}{n!}\frac{1}{L^2}(\log(Q-1) - L + \psi(n+1) - \psi(s)). \tag{11}$$

We also have double poles at $z + s = \chi_k$, $k \in \mathbb{Z}$, $k \neq 0$. Let $\varepsilon = z + s - \chi_k$, then

$$f(z) = \frac{(Q-1)^{z+s}}{(Q^{z+s}-1)^2} = (Q-1)^{\chi_k}\frac{(Q-1)^{\varepsilon}}{(Q^{\varepsilon}-1)^2}.$$

Expanding the fraction to two terms, we have

$$\frac{(Q-1)^{\varepsilon}}{(Q^{\varepsilon}-1)^2} \sim \frac{1}{\varepsilon^2 L^2}(1 + \varepsilon\log(Q-1))\left(1 - \frac{\varepsilon L}{2}\right)^2.$$

The $[N; z]$ factor ($N = n - s$) expanded to two terms around $z = \chi_k - s$ (i.e., around $\varepsilon = 0$) is

$$[n-s; \chi_k - s] \sim \frac{\Gamma(n-s+1)\Gamma(s-\chi_k)}{\Gamma(n+1-\chi_k)}[1 + \varepsilon\,\psi(n-\chi_k+1) - \varepsilon\,\psi(s-\chi_k)].$$

We put these together (including the factor $(Q-1)^{\chi_k}$) to get the residue asymptotic to ($n \to \infty$)

$$\frac{1}{L^2}\Gamma(s-\chi_k)\frac{\Gamma(n-s+1)}{\Gamma(n+1)}\mathrm{e}^{2k\pi i\log_Q n^*}(\log(Q-1) - L + \psi(n-\chi_k+1) - \psi(s-\chi_k)),$$

which holds for all $k \neq 0$, and can be summed over all $k \neq 0$ to get

$$\frac{\Gamma(n-s+1)}{\Gamma(n+1)}\frac{1}{L^2}\sum_{k\neq 0}\Gamma(s-\chi_k)\mathrm{e}^{2k\pi i\log_Q n^*} \times (\log(Q-1) - L + \psi(n-\chi_k+1) - \psi(s-\chi_k)).$$

This result can be combined with (11) to give the total residues for a typical term as

$$-\frac{2}{sL^2}(\log(Q-1) - L + \psi(n+1) - \psi(s)) - \frac{2}{s!L^2}\sum_{k\neq 0}\Gamma(s-\chi_k)\mathrm{e}^{2k\pi i\log_Q n^*}(\log(Q-1) - L + \psi(n-\chi_k+1) - \psi(s-\chi_k)).$$

Since we have $b-1$ of these terms added together, we can now sum them to get (notice that $\psi(n+1) \sim \log n$ and $\psi(n-\chi_k+1) \sim \log n$)

$$-\frac{2}{L}\left(\log_Q(Q-1) - 1 + \log_Q n\right)H_{b-1} + \frac{2}{L^2}\sum_{s=1}^{b-1}\frac{\psi(s)}{s} - \frac{2}{L}\sum_{s=1}^{b-1}\frac{1}{s!}\sum_{k\neq 0}\Gamma(s-\chi_k)\mathrm{e}^{2k\pi i\log_Q n^*}\left(\log_Q(Q-1) - 1 + \log_Q n - \frac{\psi(s-\chi_k)}{L}\right). \tag{12}$$

The known part was dealt with in the classical variance discussion (part (i)), so the total residue for terms one and three is twice (6) + (12). The $\log_Q n$ terms in the two sums on k from this expression can be rewritten as (see [6, page 174])

$$\frac{-2}{L}\log_Q n\sum_{k\neq 0}\mathrm{e}^{2k\pi i\log_Q n^*}\left[\Gamma(-\chi_k)+\sum_{s=1}^{b-1}\frac{1}{s!}\Gamma(s-\chi_k)\right]=2\log_Q n\,\delta_{E_b}(\log_Q n^*),$$

and so we have the total residue for terms one and three as

$$\begin{aligned}
&-\frac{2}{L}\left(\log_Q(Q-1)-1+\log_Q n\right)H_{b-1}\\
&\quad+\frac{2}{L^2}\sum_{s=1}^{b-1}\frac{\psi(s)}{s}+\log_Q^2(Q-1)+\frac{5}{6}+\log_Q^2 n+\frac{2\gamma}{L}\log_Q n\\
&\quad+\frac{\gamma^2}{L^2}+\frac{\pi^2}{6L^2}-2\log_Q(Q-1)+2\log_Q(Q-1)\log_Q n\\
&\quad+\frac{2\gamma}{L}\log_Q(Q-1)-2\log_Q n-\frac{2\gamma}{L}+2\log_Q n\delta_{E_b}(\log_Q n^*)\\
&\quad+\frac{2}{L}\sum_{k\neq 0}\Gamma(-\chi_k)\mathrm{e}^{2k\pi i\log_Q n^*}\left[\frac{\psi(-\chi_k)}{L}-\log_Q(Q-1)+1\right]\\
&\quad-\frac{2}{L}\sum_{s=1}^{b-1}\frac{1}{s!}\sum_{k\neq 0}\Gamma(s-\chi_k)\mathrm{e}^{2k\pi i\log_Q n^*}\left(\log_Q(Q-1)-1-\frac{\psi(s-\chi_k)}{L}\right).
\end{aligned}$$

Terms two and four of (10) can also be combined and regrouped, giving

$$\begin{aligned}
&2\sum_{0\le l<j}\left[\left(1+\cdots+\frac{(zpq^l)^{b-1}}{(b-1)!}\right)(\mathrm{e}^{z(1-pq^l-pq^j)}-\mathrm{e}^{z(1-pq^l)})\right]\quad(=:P)\\
&\quad+2\sum_{0\le l<j}\left[\mathrm{e}^{z(1-pq^l-pq^j)}\left(1+\cdots+\frac{(zpq^l)^{b-1}}{(b-1)!}\right)\left(zpq^j+\cdots+\frac{(zpq^j)^{b-1}}{(b-1)!}\right)\right]\\
&\quad(=:R)
\end{aligned}$$

Dealing with P: Let a typical term of the first bracket be

$$P_s:=2\sum_{0\le l<j}\left[\frac{(zpq^l)^s}{s!}(\mathrm{e}^{z(1-pq^l-pq^j)}-\mathrm{e}^{z(1-pq^l)})\right].$$

We simplify the expression to extract coefficients more easily:

$$\begin{aligned}
n![z^n]P_s&=2\sum_{0\le l<j}n![z^n]\frac{z^s(pq^l)^s}{s!}\sum_{k\ge 0}\left(\frac{z^k(1-pq^l-pq^j)^k}{k!}-\frac{z^k(1-pq^l)^k}{k!}\right)\\
&=2\sum_{0\le l<j}\binom{n}{s}(pq^l)^s((1-pq^l-pq^j)^{n-s}-(1-pq^l)^{n-s})\\
&\qquad(n=k+s)\quad\text{for } n \text{ large},\\
&=2\binom{n}{s}\sum_{k=0}^{n-s}\binom{n-s}{k}(-1)^k\sum_{l\ge 0}p^{s+k}q^{(s+k)l}\sum_{h\ge 1}((1+q^h)^k-1)\quad(h=j-l)
\end{aligned}$$

$$= 2\binom{n}{s}\sum_{k=0}^{N}\binom{N}{k}(-1)^k\frac{p^{s+k}}{1-q^{s+k}}\sum_{h\geq 1}((1+q^h)^k-1)$$
$$\text{(where } N := n-s).$$

We can now use Rice's method. The only poles are at $z = -s$ and $z = -s + \chi_k$ (here we use variable s instead of k). In the first case we have (see expected value)

$$f_s(z) := \frac{(Q-1)^{s+z}}{Q^{s+z}-1} \sim \frac{1}{L(s+z)},$$

making the residue $\frac{1}{L}$. The (exact) contribution of quantity $[N; z]$ around $z = -s$ was also calculated above as being $[n-s; -s] = \frac{(n-s)!(s-1)!}{n!}$, and the contribution of $H(z) = \sum_{h\geq 1}\big((1+q^h)^z - 1\big)$ is

$$H(-s) = \sum_{h\geq 1}((1+q^h)^{-s}-1) = \sum_{h\geq 1}\left(\sum_{i\geq 0}\binom{i+s-1}{i}(-q^h)^i - 1\right)$$
$$= \sum_{i\geq 1}\binom{i+s-1}{i}(-1)^i\frac{1}{Q^i-1}.$$

The total residue from the pole at $z = -s$ is thus

$$\frac{1}{L}\frac{(n-s)!(s-1)!}{n!}H(-s),$$

and by substituting this back into the expression for coefficients of P_s, and summing on s, we get

$$\sum_{s=0}^{b-1} 2\binom{n}{s}\frac{1}{L}\frac{(n-s)!(s-1)!}{n!}H(-s) = \frac{2}{L}\sum_{i\geq 1}(-1)^i\frac{1}{Q^i-1}\sum_{s=0}^{b-1}\frac{1}{i}\binom{i+s-1}{i-1}$$
$$= \frac{2}{L}\sum_{i\geq 1}\frac{(-1)^i}{i(Q^i-1)}\binom{i+b-1}{i}.$$

For the poles occurring at $z = \chi_k - s$ we have $f_s(z) \sim (Q-1)^{\chi_k}\frac{1}{\varepsilon L}$ from (8), and $[N; z]$ around $z = \chi_k - s$ is (9)

$$[n-s; \chi_k - s] \sim \Gamma(s-\chi_k)n^{\chi_k - s} \sim \Gamma(s-\chi_k)n^{\chi_k}\frac{\Gamma(n-s+1)}{\Gamma(n+1)}.$$

Again we need to calculate the contribution of the new quantity $H(z)$, which is exactly the same as before, only with $s - \chi_k$ in place of s, i.e.,

$$H(\chi_k - s) = \sum_{i\geq 1}\binom{i+s-\chi_k-1}{i}(-1)^i\frac{1}{Q^i-1},$$

to get the fluctuating residues

$$\frac{\Gamma(n-s+1)}{\Gamma(n+1)}\frac{1}{L}\sum_{k\neq 0}\mathrm{e}^{2k\pi i\log_Q n^*}\Gamma(s-\chi_k)H(\chi_k - s)$$

which can also be substituted into the expression for the coefficients of P_s and summed, so that altogether we have that the coefficients for the quantity P are

$$\frac{2}{L}\sum_{i\geq 1}\frac{(-1)^i}{i(Q^i-1)}\binom{i+b-1}{i}+\frac{2}{L}\sum_{s=0}^{b-1}\frac{1}{s!}\sum_{k\neq 0}\mathrm{e}^{2k\pi i\log_Q n^*}\Gamma(s-\chi_k)H(\chi_k-s). \tag{13}$$

Dealing with R: A typical term is

$$R_{st}:=2\sum_{0\leq l<j}\mathrm{e}^{z(1-pq^l-pq^j)}\frac{(zpq^l)^s}{s!}\frac{(zpq^j)^t}{t!},$$

with $0\leq s\leq b-1$ and $1\leq t\leq b-1$. We follow the same procedure as for P,

$$\begin{aligned}
n![z^n]R_{st} &= 2\sum_{0\leq l<j}\frac{n!}{(n-s-t)!s!t!}(1-pq^l-pq^j)^{n-s-t}(pq^l)^s(pq^j)^t\\
&\quad (n=k+s+t)\\
&= 2\frac{n!}{(n-s-t)!s!t!}\sum_{k=0}^{N}\binom{N}{k}(-1)^k\frac{p^{k+s+t}}{1-q^{k+s+t}}\sum_{h\geq 1}q^{ht}(1+q^h)^k\\
&\quad (h:=j-l,\ N:=n-s-t).
\end{aligned}$$

The residue from $\frac{(Q-1)^{z+s+t}}{Q^{z+s+t}-1}$ at $z=-s-t$ is $\frac{1}{L}$. The quantity $[N;z]$ at $z=-s-t$ is $\frac{(n-s-t)!(s+t-1)!}{n!}$, and for $H_t(z):=\sum_{h\geq 1}q^{ht}(1+q^h)^z$ we have

$$\begin{aligned}
H_t(-s-t) &= \sum_{h\geq 1}q^{ht}\sum_{i\geq 0}\binom{i+s+t-1}{i}(-q^h)^i\\
&= \sum_{i\geq 0}\binom{i+s+t-1}{i}(-1)^i\frac{1}{Q^{t+i}-1}.
\end{aligned}$$

This means that the residue from the pole at $z=-s-t$ is

$$\frac{(n-s-t)!(s+t-1)!}{Ln!}H_t(-s-t),$$

and so altogether we have

$$\begin{aligned}
&\sum_{s=0}^{b-1}\sum_{t=1}^{b-1}2\frac{n!}{(n-s-t)!s!t!}\frac{(n-s-t)!(s+t-1)!}{Ln!}H_t(-s-t)\\
&\quad =\frac{2}{L}\sum_{i\geq 0}(-1)^i\sum_{t=1}^{b-1}\frac{1}{Q^{t+i}-1}\frac{(i+t-1)!}{t!\,i!}\binom{i+b+t-1}{t+i}.
\end{aligned} \tag{14}$$

Summing the fluctuating residues from the poles at each $z=\chi_k-s-t$ gives

$$\frac{1}{L}\sum_{k\neq 0}\mathrm{e}^{2k\pi i\log_Q n^*}\Gamma(-\chi_k+s+t)\frac{\Gamma(n-s-t+1)}{\Gamma(n+1)}H_t(\chi_k-s-t),$$

so that altogether we have

$$\begin{aligned}
&\sum_{s=0}^{b-1}\sum_{t=1}^{b-1} 2\frac{n!}{(n-s-t)!s!t!}\frac{1}{L}\sum_{k\neq 0} \mathrm{e}^{2k\pi i \log_Q n^*}\Gamma(-\chi_k+s+t)n^{-s-t}H_t(\chi_k-s-t)\\
&\quad = \frac{2}{L}\sum_{s=0}^{b-1}\sum_{t=1}^{b-1}\frac{1}{s!\,t!}\sum_{k\neq 0}\mathrm{e}^{2k\pi i \log_Q n^*}\Gamma(-\chi_k+s+t)H_t(\chi_k-s-t), \qquad (15)
\end{aligned}$$

which means that the coefficients of R are (14) + (15).

All these results can be added together to get the variance in the general case. Further cancellations are dealt with in Appendix B. Finally the variance can be written as

$$\begin{aligned}
&\log_Q 2 + \frac{2}{L}\sum_{i\geq 1}\frac{(-1)^{i+b-1}}{i(Q^i-1)}\binom{i+b-1}{i}\binom{i-1}{b-1}\\
&\quad -\frac{2}{L}\sum_{j=1}^{b-1}\frac{1}{2j}\binom{2j}{j}\sum_{h\geq 0}\binom{-2j}{h}\frac{1}{Q^{h+j}-1}\\
&\quad +\frac{2}{L}\sum_{h\geq 1}\frac{(-1)^{h-1}}{h(Q^h-1)} - \frac{1}{L}\sum_{j=1}^{b-1}\frac{1}{2j}\binom{2j}{j}2^{-2j} + \delta_{V_b}(\log_Q n^*), \qquad (16)
\end{aligned}$$

where $\tilde{\delta}_{E_b}(x) = \frac{1}{L^2}\sum_{k\neq 0}\sum_{j\neq 0,\neq k}\frac{\Gamma(b-\chi_k)}{\chi_k\Gamma(b)}\frac{\Gamma(b-\chi_{k-j})}{\chi_{k-j}\Gamma(b)}\mathrm{e}^{2\pi i k x}$ in

$$\begin{aligned}
\delta_{V_b}(x) &= \frac{2}{L}\sum_{k\neq 0}\Gamma(-\chi_k)\mathrm{e}^{2k\pi i \log_Q n^*}\left[\frac{\psi(-\chi_k)}{L} - \log_Q(Q-1)+1\right]\\
&\quad + 2\delta_{E_b}(x) - \frac{2}{L}\gamma\,\delta_{E_b}(x) - \frac{2}{L}\sum_{s=1}^{b-1}\frac{1}{s!}\sum_{k\neq 0}\Gamma(s-\chi_k)\mathrm{e}^{2k\pi i \log_Q n^*}\\
&\quad \times\left(\log_Q(Q-1) - 1 - \frac{\psi(s-\chi_k)}{L}\right) - \tilde{\delta}_{E_b}(x)\\
&\quad + \frac{2}{L}\sum_{s=0}^{b-1}\frac{1}{s!}\sum_{k\neq 0}\mathrm{e}^{2k\pi i \log_Q n^*}\Gamma(s-\chi_k)H(\chi_k-s)\\
&\quad + \frac{2}{L}H_{b-1}\,\delta_{E_b}(x) - 2\log_Q(Q-1)\,\delta_{E_b}(x)\\
&\quad + \frac{2}{L}\sum_{s=0}^{b-1}\sum_{t=1}^{b-1}\frac{1}{s!\,t!}\sum_{k\neq 0}\mathrm{e}^{2k\pi i \log_Q n^*}\Gamma(-\chi_k+s+t)H_t(\chi_k-s-t).
\end{aligned}$$

This concludes the proof of Theorem 4. ■

8. The mean and variance for large *b*

To examine this variance result as $b\to\infty$, we can use results from [9] which state that

$$-\frac{1}{L}\sum_{j=1}^{b-1}\frac{1}{2j}\binom{2j}{j}2^{-2j} = -\frac{\log 2}{L} + \frac{1}{\sqrt{\pi}}b^{-\frac{1}{2}} + O(b^{-\frac{3}{2}})$$

and (for any $\varepsilon > 0$)

$$-\frac{2}{L}\sum_{j=1}^{b-1}\frac{1}{2j}\binom{2j}{j}\sum_{h\geq 0}\binom{-2j}{h}\frac{1}{Q^{h+j}-1}$$
$$=-\frac{2}{L}\sum_{m\geq 1}\log(1+Q^{-m})+O\left(\left(\frac{4}{Q(1+Q^{-1})^2}-\varepsilon\right)^b\right)$$
$$=-\frac{2}{L}\sum_{h\geq 1}\frac{(-1)^{h-1}}{h(Q^h-1)}+O\left(\left(\frac{4}{Q(1+Q^{-1})^2}-\varepsilon\right)^b\right),$$

whose big-O term is exponentially small as $b\to\infty$. By consulting [7], we can deduce that

$$\frac{2}{L}\sum_{i\geq 1}\frac{(-1)^{i+b-1}}{i(Q^i-1)}\binom{i+b-1}{i}\binom{i-1}{b-1}=O\left(\left(\frac{4}{Q(1+Q^{-1})^2}-\varepsilon\right)^b\right)$$

for any $\varepsilon > 0$ and is likewise exponentially small. Thus as $b\to\infty$, the constant in the asymptotic expansion of the variance is

$$\frac{1}{\sqrt{\pi}}b^{-\frac{1}{2}}+O(b^{-\frac{3}{2}})+O\left(\left(\frac{4}{Q(1+Q^{-1})^2}-\varepsilon\right)^b\right)+\delta_{V_b}(\log_Q n^*)=O(b^{-\frac{1}{2}}).$$

Acknowledgements

Arnold Knopfmacher's contribution was based upon work supported by the National Research Foundation under grant number 2053740. Helmut Prodinger's contribution was based upon work supported by the National Research Foundation under grant number 2053748.

Appendix A

The Fourier series (7) we want to simplify can be written as $\delta_V(x)=\sum_{k\neq 0}a_k e^{2\pi ikx}$ where

$$a_k=\frac{2}{L}\Gamma(-\chi_k)\left[\frac{\psi(-\chi_k)+\gamma}{L}-g(\chi_k)\right]-\frac{1}{L^2}\sum_{j\neq 0,\neq k}\Gamma(-\chi_j)\Gamma(-\chi_{k-j}),$$

with $g(x)=-\sum_{l\geq 1}\binom{x}{l}\frac{1}{Q^l-1}$. We consult [18] to do this, and start by using the formula $\Gamma(-x+l)(-1)^l=(x-l+1)\cdots(x-1)x\Gamma(-x)$ to rewrite

$$\Gamma(-\chi_k)g(\chi_k)=-\sum_{l\geq 1}\frac{(-1)^l\Gamma(l-\chi_k)}{l!Q^l-1},$$

so that we have

$$a_k=\frac{2}{L}\Gamma(-\chi_k)\left[\frac{\psi(-\chi_k)+\gamma}{L}\right]+\frac{2}{L}\sum_{l\geq 1}\frac{(-1)^l\Gamma(l-\chi_k)}{l!Q^l-1}$$
$$-\frac{1}{L^2}\sum_{j\neq 0,\neq k}\Gamma(-\chi_j)\Gamma(-\chi_{k-j}).$$

We now consider the function [9]

$$F(z) = L\frac{\Gamma(z)\Gamma(-\chi_k - z)}{e^{Lz} - 1},$$

with integral

$$I_1 = \frac{1}{2\pi i}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} F(z)\mathrm{d}z,$$

chosen because of the residues produced when the contour of integration is shifted. We evaluate this integral twice, by shifting the contour first left and then right. We start by shifting the line left to $\Re(z) = -\frac{1}{2}$. Simple poles occur at $z = -\chi_j$ for all $j \in \mathbb{Z} \setminus \{0\}$, with a double pole at $z = 0$.

$$\begin{aligned}
&\mathrm{Res}(F, 0) = -\gamma\Gamma(-\chi_k) - \frac{L}{2}\Gamma(-\chi_k) - \Gamma(-\chi_k)\psi(-\chi_k),\\
&\mathrm{Res}(F, -\chi_k) = -\Gamma(-\chi_k)\psi(-\chi_k) + \frac{L}{2}\Gamma(-\chi_k) - \gamma\Gamma(-\chi_k),\\
&\mathrm{Res}(F, -\chi_j) = \Gamma(-\chi_k)\Gamma(-\chi_k + \chi_j) \quad \text{for all } j \neq 0, \neq k.
\end{aligned}$$

So

$$I_1 = \frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} F(z)\mathrm{d}z - 2\Gamma(-\chi_k)(\gamma + \psi(-\chi_k)) + \sum_{j\neq 0,\neq k}\Gamma(-\chi_k)\Gamma(-\chi_k + \chi_j),$$

and we use $\frac{1}{e^{Lz}-1} = -1 - \frac{1}{e^{-Lz}-1}$ and a change of variable $z := z + \chi_k$ to get

$$2I_1 = -LI_2 - 2\Gamma(-\chi_k)(\gamma + \psi(-\chi_k)) + \sum_{j\neq 0,\neq k}\Gamma(-\chi_k)\Gamma(-\chi_k + \chi_j) \tag{17}$$

where I_2 is an integral of Mellin–Barnes type [23, p. 286ff]

$$I_2 = \frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty}\Gamma(z)\Gamma(-\chi_k - z)\mathrm{d}z = \frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty}\Gamma(z - \chi_k)\Gamma(-z)\mathrm{d}z.$$

To evaluate I_2 we shift the contour line to the right to get negative residues. The poles we consider are at $z = \chi_k$, a simple pole with residue $-\Gamma(-\chi_k)$ and at $z = l, l \in \mathbb{N}_0$, with residues $\sum_{l\geq 0}\frac{(-1)^l}{l!}\Gamma(l - \chi_k)$. So

$$\begin{aligned}
I_2 &= -\Gamma(-\chi_k) + \sum_{l\geq 0}\frac{(-1)^l}{l!}\Gamma(l - \chi_k)\\
&= -\Gamma(-\chi_k) + \Gamma(-\chi_k)\sum_{l\geq 0}\binom{\chi_k}{l} = \Gamma(-\chi_k)(e^{2\pi ik\log_Q 2} - 1).
\end{aligned}$$

On the other hand, if we write $I_1 = \frac{1}{2\pi i}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} L\frac{\Gamma(-\chi_k+z)\Gamma(-z)}{e^{Lz}-1}\mathrm{d}z$ and shift the contour of integration to the right, we collect the negative residues at $l = 1, 2, 3, \ldots$ as

$$I_1 = L \sum_{l\ge 1} \frac{(-1)^l \Gamma(l-\chi_k)}{l!Q^l - 1}. \tag{18}$$

Since we now have two expressions for I_1, which must be equal, we can combine (17) and (18), and cancel all terms except I_2, leaving us with

$$\delta_V(x) = -\frac{1}{L}\sum_{k\ne 0} \Gamma(-\chi_k)(\mathrm{e}^{2\pi ik\log_Q 2} - 1)\mathrm{e}^{2\pi ikx} = \delta_E(x + \log_Q 2) - \delta_E(x),$$

which, for $Q = 2$ is $\delta_E(x+1) - \delta_E(x)$, which is zero since $\delta_E(x)$ has period 1 [18].

Appendix B

The variance in the general case can be expressed as

$$\begin{aligned}
\mathbb{V}(d_n^{\langle b\rangle}) &= \frac{1}{12} + \frac{\pi^2}{6L^2} + \frac{2\gamma H_{b-1}}{L^2} - \frac{H_{b-1}^2}{L^2} \\
&\quad + \frac{2}{L^2}\sum_{s=1}^{b-1}\frac{\psi(s)}{s} + \frac{2}{L}\sum_{i\ge 1}\frac{(-1)^i}{i(Q^i-1)}\binom{i+b-1}{i} \\
&\quad + \frac{2}{L}\sum_{i\ge 0}(-1)^i \sum_{t=1}^{b-1}\frac{1}{Q^{t+i}-1}\frac{(i+t-1)!}{t!\,i!}\binom{i+b+t-1}{t+i} \\
&\quad - 2\log_Q(Q-1)\,\delta_{E_b}(\log_Q n^*) \\
&\quad + \frac{2}{L}\sum_{s=0}^{b-1}\sum_{t=1}^{b-1}\frac{1}{s!\,t!}\sum_{k\ne 0} e^{2\pi ik\log_Q n^*}\Gamma(-\chi_k+s+t)H_t(\chi_k-s-t) \\
&\quad + 2\delta_{E_b}(\log_Q n^*) + \frac{2}{L}\sum_{k\ne 0}\Gamma(-\chi_k)\mathrm{e}^{2\pi ik\log_Q n^*}\left[\frac{\psi(-\chi_k)}{L} - \log_Q(Q-1) + 1\right] \\
&\quad - \frac{2}{L}\gamma\,\delta_{E_b}(\log_Q n^*) - \frac{2}{L}\sum_{s=1}^{b-1}\frac{1}{s!}\sum_{k\ne 0}\Gamma(s-\chi_k)\mathrm{e}^{2\pi ik\log_Q n^*} \\
&\quad \times \left(\log_Q(Q-1) - 1 - \frac{\psi(s-\chi_k)}{L}\right) - \delta_{E_b}^2(\log_Q n^*) \\
&\quad + \frac{2}{L}\sum_{s=0}^{b-1}\frac{1}{s!}\sum_{k\ne 0}\mathrm{e}^{2\pi ik\log_Q n^*}\Gamma(s-\chi_k)H(\chi_k - s) + \frac{2}{L}H_{b-1}\,\delta_{E_b}(\log_Q n^*).
\end{aligned}$$

We can cancel terms and express

$$\begin{aligned}
&\frac{2}{L}\sum_{i\ge 1}\frac{(-1)^i}{i(Q^i-1)}\binom{i+b-1}{i} \\
&\quad + \frac{2}{L}\sum_{i\ge 0}(-1)^i\sum_{t=1}^{b-1}\frac{1}{Q^{t+i}-1}\frac{(i+t-1)!}{t!\,i!}\binom{i+b+t-1}{t+i}
\end{aligned}$$

as

$$\frac{2}{L}\sum_{t=1}^{b-1}\frac{1}{(Q^t-1)}\frac{(t-1)!}{t!}\binom{b+t-1}{t}$$
$$+\frac{2}{L}\sum_{i\ge 1}(-1)^i\sum_{t=0}^{b-1}\frac{1}{Q^{t+i}-1}\frac{(i+t-1)!}{t!\,i!}\binom{i+b+t-1}{t+i}$$
$$=\frac{2}{L}\sum_{t=1}^{b-1}\frac{1}{t(Q^t-1)}\binom{b+t-1}{t}+\frac{2}{L}\sum_{t=0}^{b-1}\sum_{i\ge t+1}\frac{(-1)^{i-t}}{Q^i-1}\frac{1}{i}\binom{i}{i-t}\binom{i+b-1}{i}$$
$$=\frac{2}{L}\sum_{t=1}^{b-1}\frac{1}{t(Q^t-1)}\binom{b+t-1}{t}$$
$$+\frac{2}{L}\sum_{t=0}^{b-1}\left[\sum_{i\ge 1}\frac{(-1)^{i-t}}{Q^i-1}\frac{1}{i}\binom{i}{i-t}\binom{i+b-1}{i}-[t\ge 1]\frac{1}{Q^t-1}\frac{1}{t}\binom{t+b-1}{t}\right]$$
$$=\frac{2}{L}\sum_{t=0}^{b-1}\sum_{i\ge 1}\frac{(-1)^{i-t}}{i(Q^i-1)}\binom{i}{i-t}\binom{i+b-1}{i}$$
$$=\frac{2}{L}\sum_{i\ge 1}\frac{(-1)^{i+b-1}}{i(Q^i-1)}\binom{i+b-1}{i}\binom{i-1}{b-1}.$$

We then use the analytic expression for harmonic numbers $H_n=\psi(n+1)+\gamma$ to rewrite

$$\frac{2}{L^2}\sum_{s=1}^{b-1}\frac{\psi(s)}{s}=\frac{2}{L^2}\sum_{s=1}^{b-1}\frac{1}{s}(H_{s-1}-\gamma)=\frac{2}{L^2}\sum_{s=1}^{b-1}\frac{1}{s}H_s-\frac{2}{L^2}\sum_{s=1}^{b-1}\frac{1}{s^2}-\frac{2\gamma}{L^2}H_{b-1}$$
$$=\frac{2}{L^2}\frac{1}{2}(H_{b-1}^2+H_{b-1}^{(2)})-\frac{2}{L^2}H_{b-1}^{(2)}-\frac{2\gamma}{L^2}H_{b-1}$$
$$=\frac{1}{L^2}H_{b-1}^2-\frac{1}{L^2}H_{b-1}^{(2)}-\frac{2\gamma}{L^2}H_{b-1}$$

which means we cancel the terms $\frac{2\gamma H_{b-1}}{L^2}-\frac{H_{b-1}^2}{L^2}$. It is also necessary to look at the term $\delta_{E_b}^2(\log_Q n^*)$, whose mean is non-zero. In [9] the square of $\delta_{E_b}(x)=-\frac{1}{L}\sum_{k\ne 0}\mathrm{e}^{2k\pi ix}\Gamma(-\chi_k)\binom{b-\chi_k-1}{b-1}$ is split into two parts — a constant (the mean of the square of the function) and the remaining periodic function of period 1 and mean zero. We write

$$\delta_{E_b}^2(x)=[\delta_{E_b}^2]_0+\tilde\delta_{E_b}(x),$$

where

$$[\delta_{E_b}^2]_0=\frac{\pi^2}{6L^2}+\frac{1}{12}-\log_Q 2+\frac{2}{L}\sum_{j=1}^{b-1}\frac{1}{2j}\binom{2j}{j}\sum_{h\ge 0}\binom{-2j}{h}\frac{1}{Q^{h+j}-1}$$
$$-\frac{2}{L}\sum_{h\ge 1}\frac{(-1)^{h-1}}{h(Q^h-1)}+\frac{1}{L}\sum_{j=1}^{b-1}\frac{1}{2j}\binom{2j}{j}2^{-2j}-\frac{H_{b-1}^{(2)}}{L^2},$$

and $\tilde\delta_{E_b}(x)$ is periodic function with mean zero. We thus have the result for the variance as in (16).

References

[1] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1970.
[2] Z.-D. Bai, H.-K. Hwang, W.-Q. Liang, Normal approximations of the number of records in geometrically distributed random variables, Random Structures Algorithms 13 (1998) 319–334.
[3] L. Devroye, A limit theory for random skip lists, Ann. Appl. Probab. 2 (1992) 597–609.
[4] P. Flajolet, G.N. Martin, Probabilistic counting algorithms for data base applications, J. Comp. Syst. Sci. 31 (1985) 182–209.
[5] P. Flajolet, R. Sedgewick, Mellin transforms and asymptotics: Finite differences and Rice's integrals, Theoret. Comput. Sci. 144 (1995) 101–124. Special volume on mathematical analysis of algorithms.
[6] R.L. Graham, D.E. Knuth, O. Patashnik, Concrete Mathematics, Addison-Wesley, 1994.
[7] P. Kirschenhofer, H. Prodinger, b-tries: A paradigm for the application of numbertheoretic methods in the analysis of algorithms, in: Contributions to General Algebra, vol. 6, Hölder–Pichler–Tempsky–Teubner, 1988, pp. 141–153.
[8] P. Kirschenhofer, H. Prodinger, On the Analysis of Probabilistic Counting, in: E. Hlawka, R.F. Tichy (Eds.), Lecture Notes in Mathematics, vol. 1452, 1990, pp. 117–120.
[9] P. Kirschenhofer, H. Prodinger, A result in order statistics related to probabilistic counting, Computing 51 (1993) 15–27.
[10] P. Kirschenhofer, H. Prodinger, W. Szpankowski, Multidimensional digital searching and some new parameters in tries, Internat. J. Found. Comput. Sci. 4 (1993) 69–84.
[11] A. Knopfmacher, H. Prodinger, Combinatorics of geometrically distributed random variables: Value and position of the rth left-to-right maximum, Discrete Math. 226 (2001) 255–267.
[12] G. Louchard, H. Prodinger, Ascending runs of sequences of geometrically distributed random variables: A probabilistic analysis, Theoret. Comput. Sci. 304 (2003) 59–86.
[13] T. Papadakis, I. Munro, P. Poblete, Average search and update costs in skip lists, BIT 32 (1992) 316–332.
[14] H. Prodinger, Combinatorial problems related to geometrically distributed random variables, in: Séminaire Lotharingien de Combinatoire (Gerolfingen, 1993), in: Prépubl. Inst. Rech. Math. Av., vol. 1993/34, Univ. Louis Pasteur, Strasbourg, 1993, pp. 87–95.
[15] H. Prodinger, Combinatorics of geometrically distributed random variables: Left-to-right maxima, Discrete Math. 153 (1996) 253–270.
[16] H. Prodinger, Combinatorics of geometrically distributed random variables: New q-tangent and q-secant numbers, Int. J. Math. Math. Sci. 24 (2000).
[17] H. Prodinger, Combinatorics of geometrically distributed random variables: Inversions and a parameter of Knuth, Ann. Comb. 5 (2001).
[18] H. Prodinger, Compositions and Patricia tries: No fluctuations in the variance! SODA (2004).
[19] W. Pugh, Skip lists: A probabilistic alternative to balanced trees, Commun. ACM 33 (1990) 668–676.
[20] R. Sedgewick, P. Flajolet, An Introduction to the Analysis of Algorithms, Addison-Wesley, 1996.
[21] W. Szpankowski, Average Case Analysis of Algorithms on Sequences, John Wiley and Sons, New York, 2001.
[22] W. Szpankowski, V. Rego, Yet another application of a binomial recurrence: Order statistics, Computing 43 (1990) 401–410.
[23] E.T. Whittaker, G.N. Watson, A Course in Modern Analysis, fourth ed., Cambridge University Press, 1927, reprinted 1973.